Well-posedness of Wasserstein Gradient Flow Solutions of Higher Order Evolution Equations

Ehsan Kamalinejad

Department of Mathematics, University of Toronto,

`ehsan@math.toronto.edu`

January 18, 2012

Abstract

A relaxed notion of displacement convexity is defined and used to establish short time existence and uniqueness of Wasserstein gradient flows for higher order energy functionals. As an application, local and global well-posedness of different higher order degenerate non-linear evolution equations are derived. Examples include the thin-film equation and the quantum drift diffusion equation in one spatial variable.

Keywords: optimal transport; Wasserstein gradient flows; displacement convexity; minimizing movement; well-posedness; thin-film equation; higher order non-linear degenerate equations.

AMS subject classification: 35A15, 35K30, 76A20, 35K65.

1 Introduction

In the last decade, the theory of gradient flows in the Wasserstein space has been a rapidly expanding area of research. With a wide range of applications to evolution equations and functional inequalities, this theory has received an extensive amount of interest. In this section we start by recalling some historical backgrounds of the theory and then we state the summary of our result. For a comprehensive discussion of all aspects of the theory we refer the reader to monographs [1] and [23].

1.1 Historical background

The Wasserstein space $\mathcal{P}_2(\mathbb{R}^m)$ consists of the Borel probability measures on $\mathbb{R}^m$ with finite second moment. The quadratic optimal transport distance, also known as the **Wasserstein**

1

distance W_2, defines a distance function between any pair of measures $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^m)$ given by

$$W_2(\mu, \nu) := \inf_\gamma \left\{ \int_{\mathbb{R}^m \times \mathbb{R}^m} |x - y|^2 \, d\gamma \ : \ \gamma \in \Gamma(\mu, \nu) \right\}^{\frac{1}{2}} \tag{1}$$

where $\Gamma(\mu, \nu) \subset \mathcal{P}_2(\mathbb{R}^m \times \mathbb{R}^m)$ is the space of probability measures with marginals μ and ν. We will refer to such measures γ as **transport plans**.

It turns out that $\mathcal{P}_2(\mathbb{R}^m)$ has a rich geometric structure and a formal Riemannian calculus can be performed on this space. The first appearance of the Riemannian calculus on $\mathcal{P}_2(\mathbb{R}^m)$ is due to Otto et al in [16] and [21]. It was shown in [21] that the solution of the porous medium equation $\partial_t u = \Delta u^m$ can be reformulated as the gradient flow of the energy $E(u) = \int \frac{u^{m+1}}{m+1}$ on the Wasserstein space. Since then, the interaction between the Riemannian space $\mathcal{P}_2(\mathbb{R}^m)$ as a geometric object and evolution equations as analytic objects have attracted a lot of attention. This point of view is commonly called "*Otto calculus*".

A notion which has been very important in the developement of this theory is the notion of *displacement convexity*. McCann in his thesis [19] introduced the notion of displacement convexity of an energy functional on the Wasserstein space. Under the displacement convexity assumption, he proved existence and uniqueness of minimizers of wide classes of energies, commonly referred to as potential, internal, and interactive energies. Displacement convexity had been defined before the development of the Wasserstein gradient flows, but after establishment of the Riemannian structure of the Wasserstein space, it turned out that displacement convexity can be interpreted as the standard convexity along the geodesics of the Wasserstein space. The displacement convexity condition, with its generalization to λ-displacement convexity, has a central role in existence, uniqueness, and long-time behaviour of the gradient flow of an energy functional.

Another important notion in the theory of Wasserstein gradient flows is the notion of *minimizing movement*. Many of the rigorous proofs of the Wasserstein gradient flows are based on the method of minimizing movement. The minimizing movement scheme was suggested by De Georgi as a variational approximation of gradient flows in general metric spaces [12]. It was later used by Jordan, Kinderlehrer, Otto [16] and by Ambrosio, Savare, Gigli [1] to construct a systematic rigorous theory of Wasserstein gradient flows. This theory was soon used by many researchers to develop existence, uniqueness, stability, long time behaviour, and numerical approximation of evolution PDEs such as in [2], [3], [6], [8], [9], [11], [13], and [20].

1.2 Summary of the results and outline of the paper

In recent years, it has become apparent that Otto calculus also applies to higher-order evolution equations, at least on a formal level. The best-studied example is the thin-film

equation $\partial_t u = -\nabla \cdot (u \nabla \Delta u)$, which corresponds to the gradient flow of the Dirichlet energy $E(u) = \frac{1}{2} \int |\nabla u|^2 \, dx$. The hope is that gradient flow methods might help to resolve long-standing problems concerning well-posedness and long-time behaviour of this PDE. However, taking advantage of the gradient flow method has proved difficult. The main obstruction has been the lack of displacement convexity of the Dirichlet energy. The same problem arises for studying other energy functionals containing derivatives of the density. In [22, open problem 5.17] Villani raised the question whether there is any useful example of a displacement convex functional that contains derivatives of the density. In [10], Carrillo and Slepčev answered this question by providing a class of displacement convex functionals. Therefore it was proved that there is no fundamental obstruction for existence of such energies. However because of the lack of displacement convexity, the Wasserstein gradient flow method has not been very successful in studying gradient flows of the Dirichlet energy and other interesting energies of higher order.

Our result can be summarized as follows:

- We introduce a relaxed notion of λ-displacement convexity of an energy functional and in Theorem 2.4 we prove that, under this relaxed assumption, the general theory of well-posedness of Wasserstein gradient flows holds at least locally.

- In Theorem 3.6, we prove that the Dirichlet energy, which is not λ-displacement convex in the standard sense, satisfies the relaxed version of λ-displacement convexity on positive measures. Hence the gradient flow of the Dirichlet energy is locally well-posed and the solution of the thin-film equation with positive initial data exists and is unique as long as positivity is preserved.

- We show that the method developed to study thin-film equation applies to a range of PDEs of higher order and different forms.

The paper is organized as follows. After recalling the backgrounds of the theory, in Section 2.2 we define the new version of λ-displacement convexity which we call *restricted λ-convexity*. Setting minor technicalities aside, the idea of restricted λ-convexity can be summarized in two simple principles: Firstly, the modulus of convexity, λ, can vary along the flow. Secondly, one can study λ-convexity *locally on sub-level sets of the energy*. Note that the local analysis of gradient flows most likely fails without the help of energy dissipation. For example, the Dirichlet energy is not even locally λ-convex, because an arbitrarily small neighbourhood of a smooth positive measure contains measures with infinite energy where λ-convexity fails altogether. Instead, by taking advantage of the defining properties of the gradient flow, we study the flow on energy sub-level sets. The key observation is that typically finiteness of the energy implies some regularity on the measure which helps to elevate the formal calculations to rigorous proofs. For example, in 1-D, densities of finite Dirichlet energy lie in H^1.

After defining restricted λ-convexity, we state our first result, Theorem 2.4. In this theorem we prove that if an energy functional is restricted λ-convex at a point μ, then the

corresponding gradient flow trajectory starting from μ exists and is unique at least for a short time. The proof is based on convergence of the minimizing movement scheme and the subdiffrential property that is carried over to the limiting curve. It is interesting that both of the constraints "locality" and "energy boundedness" are already encoded in the definition of the minimizing movement scheme (13).

In Section 3 we apply the theory developed in the previous section to the Dirichlet energy. We prove that the Dirichlet energy on $\mathbb{R}/\mathbb{Z}$, is restricted λ-convex on the measures with positive density. This theorem re-derives the existing theory [5] of well-posedness of positive solutions of the thin-film equation by a direct geometric proof. To the best of our knowledge, this is the first well-posedness result for the thin-film equation based on Wasserstein gradient flows. Two key ideas are very useful in the proofs of this section: Firstly, the Wasserstein convergence and the uniform convergence are equivalent on energy sub-level sets. Secondly, finiteness of the energy can be used directly in the calculations of the second derivative of the energy along geodesics.

In the final section, we show that the method developed in Sections 2 and 3 can be applied to a wide class of energies of different forms and of higher orders. Some important examples have been studied using this method such as equations of higher order of the form $\partial_t u = (-1)^k \partial_x (u \partial_x^{2k+1} u)$, and equations of different forms, for instance the quantum drift diffusion equation $\partial_t u = -\partial_x (u \partial_x \frac{\partial_x^2 \sqrt{u}}{\sqrt{u}})$.

The Wasserstein gradient flow approach to PDEs has some interesting features. For example, it has a unified notion of solution which allows for very weak solutions and it is applicable to equations of higher order even with the lack of maximum principle. Also the minimizing movement scheme is a constructive method. Hence the proofs are constructive and one can derive numerical approximations based on the Wasserstein gradient flows similar to what has been done in [11] and [13].

2 Well-posedness of the gradient flow

In this section, we study the well-posedness problem of gradient flows on the Wasserstein space. Informally stated, a gradient flow evolves by the steepest descent of an energy functional. This idea can be formalized in several different ways, some of which carry over to general metric spaces. Here, we consider the Fréchet subdifferential formulation of gradient flows. We will identify conditions on the energy functional that guarantee short-time existence and uniqueness. The proof is based on a careful analysis of the minimizing movement scheme.

Let us recall the notion of a gradient flow on a finite dimensional Riemannian manifold. The ingredients of a gradient flow consist of three parts: a smooth manifold M, a metric g,

and an energy E. Then the gradient flow of the energy E can be formulated as

$$\begin{cases} \partial_t x_t = V_t & \text{(velocity vector)} \\ V_t = -\nabla E(x_t) & \text{(steepest descent)}. \end{cases}$$

Note that the role of the metric is to convert the co-vector $dE = \frac{dE}{dt}dx$ into the corresponding vector ∇E on the tangent space.

In the case of Wasserstein gradient flows, the ingredients are given by: $\mathcal{P}_2(\mathbb{R}^m)$ as the manifold, the Wasserstein distance (and its infinitesimal version) as the metric, and an energy functional as the energy. The formulation of a Wasserstein gradient flow is given in (10) and as it can be seen, it is similar to its finite dimensional counterpart.

2.1   Geometry of the Wasserstein space

In this part we gather some basic elements of the Riemannian structure of the Wasserstein space $\mathcal{P}_2(\mathbb{R}^m)$. Here, we work at a formal level and we refer the reader to [1] or [22] for rigorous proofs. It is worth mentioning that we consider the Euclidean space $\mathbb{R}^m$ as the underlying space of probability measures but one can replace $\mathbb{R}^m$ with any Hilbert space by slight modifications to the definitions as it is done in [1].

Consider the Wasserstein distance (1). The Brenier-McCann theorem [7] asserts that the minimum is always assumed and the minimal transport plan is concentrated on a graph of a map $T_\mu^\nu : \mathbb{R}^m \longrightarrow \mathbb{R}^m$, provided that $\mu \in \mathcal{P}_2^a(\mathbb{R}^m)$ where $\mathcal{P}_2^a(\mathbb{R}^m)$ is the set of absolutely continuous probability measures with respect to the Lebesgue measure. In this case, we have $\nu = (T_\mu^\nu)_\# \mu$, and one can rewrite the Wasserstein distance as

$$W_2(\mu, \nu) = \left(\int_{\mathbb{R}^m} \left| T_\mu^\nu - Id \right|^2 d\mu \right)^{\frac{1}{2}}. \tag{2}$$

Assuming $\mu = u\,dx$ and $\nu = v\,dx$ are absolutely continuous measures, one can use the change of measure formula, given by the **Monge–Ampère** equation, to write an explicit relation between the densities u and v in terms of the optimal map:

$$v(T_\mu^\nu(x)) = \frac{u(x)}{det(DT_\mu^\nu)(x)}. \tag{3}$$

The optimal transport map also defines the **geodesic** μ_s between two measures μ_0 and μ_1 given by the pushforward of the linear interpolation between the optimal map $T_{\mu_0}^{\mu_1}$ and the identity map:

$$\mu_s = \left((1-s)Id + sT_{\mu_0}^{\mu_1} \right)_\# \mu_0. \tag{4}$$

The appropriate class of curves inside $\mathcal{P}_2(\mathbb{R}^m)$ which have a natural notion of tangent to them turns out to be the class of absolutely continuous curves $AC_{loc}^2([0, \infty); \mathcal{P}_2(\mathbb{R}^m))$.

A curve μ_t belongs to $AC_{loc}^2([0, \infty); \mathcal{P}_2(\mathbb{R}^m))$ if there exist a locally $L^2(dt)$ integrable function g such that

$$W_2(\mu_a, \mu_b) \leqslant \int_a^b g(t)dt \quad \forall a, b \in [0, \infty).$$

The absolutely continuous curves are given by mass conservative deformations of the measures i.e. they satisfy the **continuity equation**:

$$\partial_t \mu_t + \nabla.(\mu_t V_t) = 0$$

for a **velocity vector field** V_t of deformations of μ_t. This equation is assumed to hold in the distributional sense. In [4] Brenier and Benamou showed that $\mathcal{P}_2(\mathbb{R}^m)$ is a length space in the sense that the distance of two measures is given by the length of the shortest path between them:

$$W_2(\mu, \nu) = \inf \left\{ \int_0^1 (\int_{\mathbb{R}^m} |V_t|^2 d\mu_t)^{\frac{1}{2}} dt \ \text{ s.t. } \ \partial_t \mu_t + \nabla.(\mu_t V_t) = 0; \ \mu_0 = \mu, \ \mu_1 = \nu \right\} \quad (5)$$

where the infimum is taken over all curves in $AC^2([0, 1]; \mathcal{P}_2(\mathbb{R}^m))$. For a given curve $\mu_t \in AC^2([0, 1]); \mathcal{P}_2^a(\mathbb{R}^m))$ there might be many velocity vectors that satisfy the same continuity equation $\partial_t \mu_t + \nabla.(\mu_t V_t) = 0$. For instance vector fields of the form $F_t + V_t$ where $\nabla.(\mu_t F_t) = 0$ all satisfy the same continuity equation. However there is a unique vector field that minimizes (5), i.e. the one which defines the distance between μ and ν. This optimal velocity vector field is defined to be the **tangent vector field** to the curve μ_t. The tangent vector field of μ_t can also be expressed in term of the optimal maps. If V_t is the tangent vector field of μ_t then

$$V_t = \lim_{\epsilon \to 0} \frac{T_{\mu_t}^{\mu_{t+\epsilon}} - Id}{\epsilon}.$$

The converse is also true, i.e. for a given optimal map T_μ^ν, the vector field $T_\mu^\nu - Id$ is a tangent vector at μ for some curve that passes μ. The tangent vector fields are also useful in calculating the derivative of the Wasserstein metric along curves. Let $\mu_t \in AC_{loc}^2(\mathbb{R}^+; \mathcal{P}_2^a(\mathbb{R}^m))$. By [1, Chapter 8] **the derivative of the Wasserstein metric** along the curve μ_t is given by

$$\frac{d}{dt} W_2(\mu_t, \nu)^2 = 2 \int_{\mathbb{R}^m} \langle V_t, Id - T_{\mu_t}^\nu \rangle d\mu_t \quad \forall \nu \in \mathcal{P}_2(\mathbb{R}^m) \quad (6)$$

where V_t is the tangent vector field to μ_t and $\langle ., . \rangle$ is the standard inner product on $\mathbb{R}^m$.

The Wasserstein metric is closely related to a certain weak topology on $\mathcal{P}_2(\mathbb{R}^m)$, induced by **narrow convergence**:

$$\mu_n \xrightarrow{\text{narrow}} \mu \iff \int_{\mathbb{R}^m} f d\mu_n \to \int_{\mathbb{R}^m} f d\mu \quad \forall f \in C_b^0(\mathbb{R}^m), \quad (7)$$

where $C_b^0(\mathbb{R}^m)$ is the set of of continuous bounded real functions on $\mathbb{R}^m$. The topologies induced by the narrow convergence and the Wasserstein distance are equivalent for sequences

of measures with uniformly bounded second moments:

$$\lim_{n\to\infty} W_2(\mu_n, \mu) = 0 \iff \begin{cases} \mu_n \xrightarrow{\text{narrow}} \mu \\ \{\mu_n\} \text{ has uniformly bounded 2-moments.} \end{cases} \tag{8}$$

2.2 Wasserstein gradient flows

Now we describe gradient flows on the Wasserstein space. Consider the energy functional $E : \mathcal{P}_2(\mathbb{R}^m) \to [0, \infty]$ and let its domain, $D(E)$, be the set where E is finite. Let μ lie in $D(E) \cap \mathcal{P}_2^a(\mathbb{R}^m)$. A vector field $\xi \in L^2(d\mu)$ belongs to the **subdifferential** of E at μ if

$$\liminf_{\substack{\nu \to \mu \\ \nu \in D(E)}} \frac{E(\nu) - E(\mu) - \int_X \langle \xi, T_\mu^\nu - Id \rangle \, d\mu}{W_2(\mu, \nu)} \geqslant 0. \tag{9}$$

We say that a curve $\mu_t \in AC_{loc}^2(\mathbb{R}^+, \mathcal{P}_2^a(\mathbb{R}^m))$ is a trajectory of the **gradient flow** for the energy E, if there exists a velocity field V_t with $|V_t|_{L^2(d\mu_t)} \in L_{loc}^1(\mathbb{R}^+)$ such that

$$\begin{cases} \partial_t \mu_t + \nabla \cdot (\mu_t V_t) = 0 & \text{(continuity equation),} \\ V_t \in -\partial E(\mu_t) & \text{(steepest descent)} \end{cases} \tag{10}$$

hold for almost every $t > 0$. We will refer to (10) as the **gradient flow equation**. The continuity equation, which is assume to hold in the distributional sense, links the curve with its velocity vector field and ensures that the mass is conserved. The steepest descent equation expresses that the gradient flow evolves in the direction of maximal energy dissipation.

Next we describe the link between Wasserstein gradient flows and evolution PDEs. Let $\mu = udx$ be in $D(E)$ and let $V \in \partial E(\mu)$ be a tangent vector field at μ. Consider a linear perturbation of μ given by the curve $\mu_\epsilon := (Id + \epsilon W)_\# \mu$ for small values of $\epsilon > 0$ where $W \in C_c^\infty(\mathbb{R}^m; \mathbb{R}^m)$. By the subdifferential inequality (9) we have

$$\limsup_{\epsilon \uparrow 0} \frac{E(u_\epsilon) - E(u)}{\epsilon} \leqslant \int_{\mathbb{R}^m} \langle V, W \rangle udx \leqslant \liminf_{\epsilon \downarrow 0} \frac{E(u_\epsilon) - E(u)}{\epsilon}.$$

On the other hand, assuming C^2 regularity on u_ϵ and using standard first order variations we have

$$\lim_{\epsilon \to 0} \frac{E(u_\epsilon) - E(u)}{\epsilon} = \int_{\mathbb{R}^m} (\frac{\delta E(u)}{\delta u})(\frac{\partial u_\epsilon}{\partial \epsilon}) udx$$

where $\dfrac{\delta E(u)}{\delta u}$ stands for standard first variations of E. Therefore

$$\int_{\mathbb{R}^m} \langle V, W \rangle udx = \int_{\mathbb{R}^m} \frac{\delta E(u)}{\delta u} \partial_\epsilon u_\epsilon dx.$$

The continuity equation for the curve u_ϵ implies that $\partial_\epsilon u_\epsilon = -\nabla.(uW)$. Hence

$$\int_{\mathbb{R}^m} \langle V, W \rangle u \, dx = - \int_{\mathbb{R}^m} \{ \frac{\delta E(u)}{\delta u} \nabla.(uW) \} dx.$$

Integrating by parts, we have

$$\int_{\mathbb{R}^m} \langle V, W \rangle u \, dx = \int_{\mathbb{R}^m} \langle \nabla(\frac{\delta E(u)}{\delta u}), W \rangle u \, dx.$$

Since W is arbitrary we have

$$V(x) = \nabla(\frac{\delta E(u)}{\delta u})(x) \quad \text{for } \mu\text{-a.e. } x. \tag{11}$$

Now assume that a curve $\mu_t = u_t dx$ satisfies the gradient flow equation (10). Steepest descent equation and (11) imply

$$V_t(x) = -\nabla(\frac{\delta E(u)}{\delta u})(x).$$

By plugging V_t into continuity equation, we have

$$\partial_t u = \nabla. \left(u \nabla(\frac{\delta E(u)}{\delta u}) \right). \tag{12}$$

This is the corresponding PDE for the gradient flow of the energy E. For example in the case of the Dirichlet energy $E(u) = \int_{\mathbb{R}^m} |\nabla u|^2 dx$, the first variation is given by $\frac{\delta E(u)}{\delta u} = -\Delta u$. Therefore the corresponding PDE is the **thin-film equation**:

$$\partial_t u = -\nabla.(u \nabla \Delta u).$$

Our proofs are based on the **minimizing movement scheme** as a discrete-time approximation of a gradient flow which is described here. Let $\mu_0 \in D(E)$, and fix the step size $\tau > 0$. Recursively define a sequence $\{M_\tau^n\}_{n=1}^{+\infty}$ by setting $M_0^\tau = \mu_0$, and for $n \geqslant 1$,

$$M_n^\tau = \operatorname*{argmin}_{\mu \in D(E)} \left\{ E(\mu) + \frac{1}{2\tau} W_2^2 \left(M_{n-1}^\tau, \mu \right) \right\}. \tag{13}$$

The formal Euler-Lagrange equation for this minimization problem is given by

$$U_n^\tau \in -\partial E \left(M_n^\tau \right)$$

where $U_n^\tau = -\frac{T_{M_n^\tau}^{M_{n-1}^\tau} - Id}{\tau}$. Next we define a piecewise constant curve and a corresponding velocity field by

$$\mu_t^\tau := M_n^\tau, \qquad V_t^\tau := U_n^\tau, \qquad \text{for } (n-1)\tau < t \leq n\tau.$$

8

We have

$$V_t^\tau \in -\partial E(\mu_t^\tau) \quad \forall t > 0. \tag{14}$$

This equation suggests that μ_t^τ is an approximation of the gradient flow trajectory of E starting from μ_0.

There is a standard set of hypothesises that we assume throughout this section. We gather the hypothesises here:

- E is nonnegative, and its sub-level sets are locally compact in the Wasserstein space.

- E is lower semicontinuous under narrow convergence.

- $D(E) \subseteq \mathcal{P}_2^a(\mathbb{R}^m)$.

The first two conditions guarantee existence and convergence of minimizing movement scheme (13). The third condition ensures that measures of finite energy are absolutely continuous, allowing us to use transport maps for studying the Wasserstein distance which simplifies the calculations, and allow us to view the subdifferential as a tangent vector. Note that these conditions can be sharpened as in [1], but to make the presentation more apparent, we prefer to work in this more concrete setting.

The main condition that guarantees existence and uniqueness of a Wasserstein gradient flow is given by the displacement convexity condition which asks for convexity of the energy along geodesics of the Wasserstein space. Let $\mu_s : [0, 1] \longrightarrow \mathcal{P}_2(\mathbb{R}^m)$ be the geodesic between $\mu_0, \mu_1 \in \mathcal{P}_2(X)$. An energy functional E is called **displacement convex** along μ_s if

$$E(\mu_s) \leqslant (1 - s)E(\mu_0) + sE(\mu_1) \quad s \in [0, 1].$$

More generally the energy is called λ-**displacement convex** or in short λ-convex if the convexity is bounded from below by the constant λ, i.e.

$$E(\mu_s) \leqslant (1 - s)E(\mu_0) + sE(\mu_1) - \frac{\lambda}{2}s(1 - s)W_2^2(\mu_0, \mu_1) \quad s \in [0, 1]. \tag{15}$$

Furthermore, assuming that $E(\mu_s)$ is smooth as a function of s, one can write a derivative version of λ-convexity. In this case, E is λ-convex along μ_s if

$$\frac{d^2}{ds^2}E(\mu_s) \geqslant \lambda W_2^2(\mu_0, \mu_1). \tag{16}$$

An energy is called λ-convex if it satisfies (15) along all geodesics of $\mathcal{P}_2(\mathbb{R}^m)$.
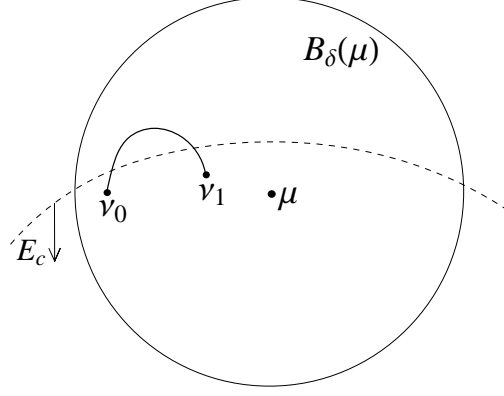
Figure 1: Restricted λ-convexity

2.3 Restricted λ-convexity and local well-posedness

Definition 2.1 (Restricted λ-convexity.) *We say that an energy E is restricted λ-convex at $\mu \in D(E)$ with $E(\mu) < c < +\infty$, if $\exists \delta > 0$, such that E is λ-convex along the geodesics connecting any pair of measures $\nu_1, \nu_2 \in B_\delta(\mu) \cap E_c$, where $B_\delta(\mu) = \{\nu \quad s.t. \quad W_2(\mu, \nu) < \delta\}$ and $E_c = \{\nu \quad s.t. \quad E(\nu) < c\}$.*

The following lemma has a key role in the arguments of Theorem 2.4. It shows how one can use restricted λ-convexity assumption to study the subdifferential of an energy.

Lemma 2.2 (Subdifferential and restricted λ-convexity.) *Assume that E is restricted λ-convex at μ. Then a vector field $\xi \in L^2(d\mu)$ belongs to the subdifferential of E at μ if and only if*

$$E(\nu) - E(\mu) \geqslant \int_X \langle \xi, T_\mu^\nu - Id \rangle \, \mathrm{d}\mu + \frac{\lambda}{2} W_2^2(\mu, \nu) \quad \forall \nu \in B_\delta(\mu) \cap E_c \qquad (17)$$

where $B_\delta(\mu) \cap E_c$ is the corresponding restricted λ-convexity domain at μ.

Proof. First we claim that for studying the subdifferential of the functional, it is enough to consider the restricted domain $B_\delta(\mu) \cap E_c$. Let $\xi \in L^2(d\mu)$, we have to show that

$$\liminf_{\substack{\nu \to \mu \\ \nu \in D(E)}} \frac{E(\nu) - E(\mu) - \int_X \langle \xi, T_\mu^\nu - Id \rangle \, \mathrm{d}\mu}{W_2(\mu, \nu)} \geqslant 0. \qquad (18)$$

if and only if

$$\liminf_{\substack{\nu \to \mu \\ \nu \in B_\delta(\mu) \cap E_c}} \frac{E(\nu) - E(\mu) - \int_X \langle \xi, T_\mu^\nu - Id \rangle \, \mathrm{d}\mu}{W_2(\mu, \nu)} \geqslant 0. \qquad (19)$$

$\Downarrow$ is trivial.
For $\Uparrow$ assume that $\{\nu_n\}_1^\infty$ is a minimizing sequence for (18).

10

In the case that $\liminf_{n\to\infty} E(\nu_n) - E(\mu) > 0$ we have

$$\frac{E(\nu_n) - E(\mu) - \int_X \langle \xi, T_\mu^{\nu_n} - Id \rangle \, d\mu}{W_2(\mu, \nu_n)} \geqslant \frac{E(\nu_n) - E(\mu) - \left(\int_X |\xi|^2 \, d\mu\right)^{1/2} \left(\int_X |T_\mu^{\nu_n} - Id|^2 \, d\mu\right)^{1/2}}{W_2(\mu, \nu_n)}$$

$$= \frac{E(\nu) - E(\mu)}{W_2(\mu, \nu_n)} - \left(\int_X |\xi|^2 \, d\mu\right)^{1/2} \frac{\left(\int_X |T_\mu^{\nu_n} - Id|^2 \, d\mu\right)^{1/2}}{W_2(\mu, \nu_n)}$$

$$= \left[\frac{E(\nu) - E(\mu)}{W_2(\mu, \nu_n)} - |\xi|_{L_\mu^2}\right] \longrightarrow +\infty.$$

Therefore inequality (18) is automatically true if $\liminf_{n\to\infty} E(\nu_n) - E(\mu) > 0$. Hence one only needs to consider sequences $\{\nu_n\}_1^\infty$ such that $\lim_{n\to\infty} E(\nu_n) - E(\mu) \leqslant 0$. Therefore, for large enough n we have $E(\nu_n) < c$. On the other hand, $\nu_n \xrightarrow{W_2} \mu$. Hence (19) and (18) are equivalent.

It is clear that (17) implies (19). Conversely, let $\xi \in L^2(d\mu)$ satisfy (19). Let $\nu \in B_\delta(\mu) \cap E_c$. Since E is restricted λ-convex at μ, we have λ-convexity of E along the geodesic μ_s connecting μ to ν. Therefore

$$E(\mu_s) \leqslant (1-s)E(\mu) + sE(\nu) - \frac{\lambda}{2}s(1-s)W_2^2(\mu, \nu) \quad \forall s \in [0,1].$$

Dividing by s and reordering, we have

$$\frac{E(\mu_s) - E(\mu)}{s} \leqslant E(\nu) - E(\mu) - \frac{\lambda}{2}(1-s)W_2^2(\mu, \nu). \tag{20}$$

ξ is in the subdifferential of E at μ. Hence

$$\liminf_{s\to 0^+} \frac{E(\mu_s) - E(\mu)}{s} \geqslant \lim_{s\to 0^+} \frac{1}{s} \int_X \langle \xi, T_\mu^{\mu_s} - Id \rangle \, d\mu$$

$$= \int_X \langle \xi, T_\mu^\nu - Id \rangle \, d\mu \tag{21}$$

where we used linearity of the interpolate map $T_\mu^{\mu_s} = Id + s(T_\mu^\nu - Id)$. Therefore (20) and (21) imply

$$E(\nu) - E(\mu) \geqslant \int_X \langle \xi, T_\mu^\nu - Id \rangle \, d\mu + \frac{\lambda}{2}W_2^2(\mu, \nu).$$

$\square$

The following lemma is used in Theorem 2.4 when we study weak convergence of tangent vector fields.

Lemma 2.3 *Let $\mu_t^k, \mu_t \in AC^2([0, \hat{t}]; \mathcal{P}_2^a(\mathbb{R}^m))$ and let $V_t^k \in L^2(d\mu_t^k)$, $V_t \in L^2(d\mu_t)$. Assume that*

- $\mu_t^k \xrightarrow{W_2} \mu_t$ *uniformly on $[0, \hat{t}]$.*

- V_t^k ***weakly converges*** *to V_t in the sense that $\forall U \in C_b^0([0, \hat{t}] \times \mathbb{R}^m)$, we have*

$$\lim_{k \to \infty} \int_0^{\hat{t}} \int_{\mathbb{R}^m} \langle V_t^k, U(t, x) \rangle \, d\mu_t^k dt = \int_0^{\hat{t}} \int_{\mathbb{R}^m} \langle V_t, U(t, x) \rangle \, d\mu_t dt.$$

Then $\forall \nu \in \mathcal{P}_2(\mathbb{R}^m)$

$$\lim_{k \to \infty} \int_0^{\hat{t}} \int_{\mathbb{R}^m} \left\langle V_t^k, T_{\mu_t^k}^\nu - Id \right\rangle d\mu_t^k dt = \int_0^{\hat{t}} \int_{\mathbb{R}^m} \left\langle V_t, T_{\mu_t}^\nu - Id \right\rangle d\mu_t dt.$$

Proof. Since V_t^k is weakly convergent by uniform boundedness principle we have $\sup_k \int_0^{\hat{t}} \int_{\mathbb{R}^m} |V_t^k|^2 d\mu_t^k dt < +\infty$. Let

$$M = \sup_k \int_0^{\hat{t}} \left(\int_{\mathbb{R}^m} \left| V_t^k \right|^2 d\mu_t^k + \int_{\mathbb{R}^m} |V_t|^2 \, d\mu_t \right) dt.$$

Choose $T_t \in C_c^0([0, \hat{t}] \times \mathbb{R}^m)$ such that

$$\int_0^{\hat{t}} \int_{\mathbb{R}^m} |T_{\mu_t}^\nu - T_t|^2 d\mu_t dt < \epsilon^2.$$

We have

$$\left| \int_0^{\hat{t}} \int_{\mathbb{R}^m} \left\langle V_t^k, T_{\mu_t^k}^\nu - Id \right\rangle d\mu_t^k dt - \int_0^{\hat{t}} \int_{\mathbb{R}^m} \left\langle V_t, T_{\mu_t}^\nu - Id \right\rangle d\mu_t dt \right|$$

$$\leqslant \underbrace{\left| \int_0^{\hat{t}} \int_{\mathbb{R}^m} \left\langle V_t^k, Id \right\rangle d\mu_t^k dt - \int_0^{\hat{t}} \int_{\mathbb{R}^m} \left\langle V_t, Id \right\rangle d\mu_t dt \right|}_{A}$$

$$+ \underbrace{\left| \int_0^{\hat{t}} \int_{\mathbb{R}^m} \left\langle V_t^k, T_{\mu_t^k}^\nu - T_t \right\rangle d\mu_t^k dt \right|}_{B}$$

$$+ \underbrace{\left| \int_0^{\hat{t}} \int_{\mathbb{R}^m} \left\langle V_t^k, T_t \right\rangle d\mu_t^k dt - \int_0^{\hat{t}} \int_{\mathbb{R}^m} \left\langle V_t, T_{\mu_t}^\nu \right\rangle d\mu_t dt \right|}_{C}$$

We study each of the items separately.

Since μ_t^k is uniformly converging to μ_t, the second moment of μ_t^k is uniformly bounded. In particular there is a compact set $S \subset [0, \hat{t}] \times \mathbb{R}^m$ such that

$$\left(\int_{S^c} |x|^2 \mu_t dt + \sup_k \int_{S^c} |x|^2 \mu_t^k dt \right) < \epsilon^2. \tag{22}$$

12

We have

$$\lim_{k\to\infty} A = \lim_{k\to\infty} \left| \int_0^{\hat{t}} \int_{\mathbb{R}^m} \langle V_t^k, Id \rangle \, d\mu_t^k dt - \int_0^{\hat{t}} \int_{\mathbb{R}^m} \langle V_t, Id \rangle \, d\mu_t dt \right|$$

$$\leqslant \lim_{k\to\infty} \left| \int_S \langle V_t^k, Id \rangle \, d\mu_t^k dt - \int_S \langle V_t, Id \rangle \, d\mu_t dt \right|$$

$$+ \lim_{k\to\infty} \left| \int_{S^c} \langle V_t^k, Id \rangle \, d\mu_t^k dt - \int_{S^c} \langle V_t, Id \rangle \, d\mu_t dt \right|.$$

Because S is compact one can use weak convergence of V_t^k on S. Hence the limit of the first term vanishes and we have

$$\lim_{k\to\infty} A \leqslant \lim_{k\to\infty} \left| \int_{S^c} \langle V_t^k, Id \rangle \, d\mu_t^k dt - \int_{S^c} \langle V_t, Id \rangle \, d\mu_t dt \right|$$

$$\leqslant \frac{\epsilon}{2} \lim_{k\to\infty} \left(\int_{S^c} |V_t^k|^2 d\mu_t^k dt + \int_{S^c} |V_t|^2 d\mu_t dt \right)$$

$$+ \lim_{k\to\infty} \frac{1}{2\epsilon} \left(\int_{S^c} |x|^2 d\mu_t^k dt + \int_{S^c} |x|^2 d\mu_t dt \right)$$

where we used Young's inequality with the constant ϵ. By (22) we have

$$\lim_{k\to\infty} A \leqslant \frac{\epsilon M}{2} + \frac{\epsilon}{2}.$$

Since ϵ is arbitrary we have $\lim_{k\to\infty} A = 0$.

We now study B. Consider the measure γ_t^k on $\mathbb{R}^m \times \mathbb{R}^m$ given by

$$\gamma_t^k = (Id \times T_{\mu_t^k}^\nu)_\# \mu_t^k.$$

Recall that the measure γ_t^k is the optimal plan with marginals μ_t^k and ν. Since $\mu_t^k \to \mu_t$, by the stability of optimal plans [23, Theorem 5.20], the set of optimal plans between μ_t^k and ν is compact in the narrow topology and every limit point is an optimal plan between μ_t and ν. On the other hand, because μ_t is an absolutely continuous measure, Brenier-McCann Theorem ensures that the optimal plan between μ_t and ν is unique. This implies that the sequence γ_t^k converges narrowly for all $t \in [0, \hat{t}]$. Furthermore, the uniform convergence of μ_t^k implies that γ_t^k have uniformly bounded second moment. We have

$$\gamma_t^k = (Id \times T_{\mu_t^k}^\nu))_\# \mu_t^k \xrightarrow{narrow} \gamma_t = (Id \times T_{\mu_t}^\nu))_\# \mu_t \quad \forall t \in [0, \hat{t}]. \tag{23}$$

Taking the limit of B yields

$$\lim_{k\to\infty} B = \lim_{k\to\infty} \left| \int_0^{\hat{t}} \int_{\mathbb{R}^m} \left\langle V_t^k, T_{\mu_t^k}^\nu - T_t \right\rangle d\mu_t^k dt \right|$$

$$\leqslant \lim_{k\to\infty} \frac{\epsilon}{2} \int_0^{\hat{t}} \int_{\mathbb{R}^m} |V_t^k|^2 d\mu_t^k dt + \frac{1}{2\epsilon} \int_0^{\hat{t}} \int_{\mathbb{R}^m} |T_{\mu_t^k}^\nu - T_t| d\mu_t^k dt$$

$$\leqslant \frac{\epsilon M}{2} + \frac{1}{2\epsilon} \lim_{k\to\infty} \int_0^{\hat{t}} \int_{\mathbb{R}^m} |T_{\mu_t^k}^\nu - T_t|^2 d\mu_t^k dt.$$

13

by lifting to the optimal plans $\gamma_t^k = (Id \times T_{\mu_t^k}^\nu)_\# \mu_t^k$, we have

$$\lim_{k \to \infty} B \leqslant \frac{\epsilon M}{2} + \frac{1}{2\epsilon} \lim_{k \to \infty} \int_0^{\hat{t}} \int_{\mathbb{R}^m \times \mathbb{R}^m} |y - T_t(x)|^2 d\gamma_t^k dt.$$

Since $\gamma_t^k \to \gamma_t$ point-wise, γ_t^k has uniformly bounded second moment, and $|y - T_t(x)|^2$ is dominated by a constant times $|x^2 + y^2 + 1|$, we can use dominated convergence theorem. Therefore

$$\lim_{k \to \infty} B \leqslant \frac{\epsilon M}{2} + \frac{1}{2\epsilon} \lim_{k \to \infty} \int_0^{\hat{t}} \int_{\mathbb{R}^m \times \mathbb{R}^m} |y - T_t(x)|^2 d\gamma_t^k dt$$

$$= \frac{\epsilon M}{2} + \frac{1}{2\epsilon} \int_0^{\hat{t}} \int_{\mathbb{R}^m \times \mathbb{R}^m} |y - T_t(x)|^2 d\gamma_t dt$$

$$= \frac{\epsilon M}{2} + \frac{1}{2\epsilon} \int_0^{\hat{t}} \int_{\mathbb{R}^m} |T_{\mu_t}^\nu - T_t|^2 d\mu_t dt$$

$$\leqslant \frac{\epsilon M}{2} + \frac{\epsilon}{2}.$$

Finally, we study the last term C. We have

$$\lim_{k \to \infty} C = \lim_{k \to \infty} \left| \int_0^{\hat{t}} \int_{\mathbb{R}^m} \langle V_t^k, T_t \rangle d\mu_t^k dt - \int_0^{\hat{t}} \int_{\mathbb{R}^m} \langle V_t, T_{\mu_t}^\nu \rangle d\mu_t dt \right|$$

$$\leqslant \lim_{k \to \infty} \left| \int_0^{\hat{t}} \int_{\mathbb{R}^m} \langle V_t^k, T_t \rangle d\mu_t^k dt - \int_0^{\hat{t}} \int_{\mathbb{R}^m} \langle V_t, T_t \rangle d\mu_t dt \right|$$

$$+ \left| \int_0^{\hat{t}} \int_{\mathbb{R}^m} \langle V_t, T_{\mu_t}^\nu - T_t \rangle d\mu_t dt \right|.$$

Since $T_t \in C_b^0$ we can use weak convergence of V_t^k for the first term. Hence

$$\lim_{k \to \infty} C \leqslant \left| \int_0^{\hat{t}} \int_{\mathbb{R}^m} \langle V_t, T_{\mu_t}^\nu - T_t \rangle d\mu_t dt \right|$$

$$\leqslant M \int_0^{\hat{t}} \int_{\mathbb{R}^m} |T_{\mu_t}^\nu - T_t|^2 d\mu_t dt$$

$$\leqslant M\epsilon^2.$$

$\square$

Theorem 2.4 (Existence and uniqueness of the flow) *Let $E : \mathcal{P}_2(\mathbb{R}^m) \longrightarrow [0, +\infty]$ be a lower semi continuous energy functional with locally compact sub-level sets and let $D(E) \subseteq \mathcal{P}_2^a(\mathbb{R}^m)$. Assume $E(\mu) < c < +\infty$ and that E is restricted λ-convex at μ. Then there exist $\hat{t} > 0$ and a curve $\mu_t \in AC^2\left([0, \hat{t}]; \mathcal{P}_2^a(\mathbb{R}^m)\right)$ such that μ_t is the unique gradient flow of E starting from μ.*

Proof. Let $\mu_t^k := \mu_t^\tau$ be a piecewise constant solution to the minimizing movement scheme (13) with $\tau = \frac{1}{k}$. The minimizing movement sequence is designed in a way that it converges to a limiting curve in a very general setting. In [1, Theorem 11.1.6] it has been proved that, under very weak assumptions which hold here, the minimizing movement scheme converges sub-sequentially to a limiting curve such that (after relabelling) $\forall a > 0$

- $\mu_t^k \xrightarrow{W_2} \mu_t \in AC^2([0, a]; \mathcal{P}_2(\mathbb{R}^m))$ uniformly in $[0, a]$.

- The sequence $\{V_t^k\}$ of the velocity tangent vectors to $\{\mu_t^k\}$ converges weakly to $V_t \in L^2(d\mu)$ in $\mathbb{R}^m \times (0, T)$.

- The continuity equation $\partial_t \mu_t + \nabla \cdot (\mu_t V_t) = 0$ holds for the limiting curve.

We need to prove that the limiting curve μ_t satisfies the steepest descent equation and that it is unique. Since μ_t is a continuous curve, we can find $\hat{t}$ such that $\mu_t \in B_{\delta/4}(\mu)$ for all $t \in [0, \hat{t}]$ where δ is the radius of restricted λ-convexity at μ. We have

$$E(\mu_t) \leqslant \liminf_{k \to \infty} E(\mu_t^k) \leqslant E(\mu) < c. \tag{24}$$

The first inequality follows from lower semi continuity of the energy and the second inequality follows from the structure of the minimizing movement scheme (13). Hence

$$\mu_t \in B_{\delta/4}(\mu) \cap E_c \quad \forall t \in [0, \hat{t}]. \tag{25}$$

Since $\mu_t^k \xrightarrow{W_2} \mu_t$ uniformly in $[0, \hat{t}]$, we can find $K \in \mathbb{N}$ such that $W_2(\mu_t, \mu_t^k) < \delta/4$, $\forall k \geqslant K$ and $\forall t \in [0, \hat{t}]$. Without loss of generality we assume that $K = 1$. Therefore (24) and (25) imply

$$\mu_t^k \in B_{\delta/2}(\mu) \cap E_c. \tag{26}$$

The Euler-Lagrange equation of minimizing movement (14) implies that $-V_t^k \in \partial E(\mu_t^k)$. Therefore using (26) and the variational formulation of the subdifferential (Lemma 2.2) we have

$$E(\nu) - E(\mu_t^k) \geqslant \int_{\mathbb{R}^m} \left\langle -V_t^k, T_{\mu_t^k}^\nu - Id \right\rangle d\mu_t^k + \frac{\lambda}{2} W_2^2(\mu_t^k, \nu) \tag{27}$$

for all $\nu \in B_{\delta/2}(\mu) \cap E_c$. By construction, $B_{\delta/4}(\mu_t) \cap E_c \subseteq B_{\delta/2}(\mu_t^k) \cap E_c$. Therefore (27) holds for all ν in $B_{\delta/4}(\mu_t) \cap E_c$. By integrating (27) over t and against a text function $\psi \in C_c^\infty((0, \hat{t}); [0, \infty))$ we have

$$\int_0^{\hat{t}} E(\nu)\psi(t)dt - \int_0^{\hat{t}} E(\mu_t^k)\psi(t)dt \geqslant \int_0^{\hat{t}} \int_{\mathbb{R}^m} \left\langle -V_t^k, T_{\mu_t^k}^\nu - Id \right\rangle \psi(t)d\mu_t^k dt + \frac{\lambda}{2} \int_0^{\hat{t}} W_2^2(\mu_t^k, \nu)\psi(t)dt \tag{28}$$

for all ν in $B_{\delta/4}(\mu_t) \cap E_c$.

We take the limit of (28) as $k \to \infty$. By the lower semi-continuity of E

$$\int_0^{\hat{t}} E(\nu)\psi(t)dt - \int_0^{\hat{t}} E(\mu_t)\psi(t)dt \geqslant \int_0^{\hat{t}} E(\nu)\psi(t)dt - \liminf_{k \to \infty} \int_0^{\hat{t}} E(\mu_t^k)\psi(t)dt.$$

15

Lemma 2.3 implies that

$$\lim_{k\to\infty} \int_0^{\hat{t}} \int_{\mathbb{R}^m} \left\langle V_t^k, T_{\mu_t^k}^\nu - Id \right\rangle \psi(t) d\mu_t^k dt = \int_0^{\hat{t}} \int_{\mathbb{R}^m} \left\langle V_t, T_{\mu_t}^\nu - Id \right\rangle \psi(t) d\mu_t dt.$$

By the triangle inequality

$$W_2(\mu_t, \nu) - W_2(\mu_t^k, \mu_t) \leqslant W_2(\mu_t^k, \nu) \leqslant W_2(\mu_t, \nu) + W_2(\mu_t^k, \mu_t). \tag{29}$$

Therefore $\lim_{k\to\infty} W_2(\mu_t^k, \nu) = W_2(\mu_t, \nu)$. Furthermore, since $\mu_t^k \xrightarrow{W_2} \mu_t$ uniformly, inequality (29) implies that $W_2(\mu_t^k, \nu)$ is uniformly bounded. Hence, by dominated convergence theorem

$$\lim_{k\to\infty} \int_0^{\hat{t}} W_2^2(\mu_t^k, \nu)\psi(t) dt = \int_0^{\hat{t}} W_2^2(\mu_t, \nu)\psi(t) dt.$$

In conclusion $\forall \nu \in B_{\delta/4}(\mu_t) \cap E_c$ and $\forall \psi \in C_c^\infty((0, \hat{t}); [0, \infty))$ we have

$$\int_0^{\hat{t}} E(\nu)\psi(t) dt - \int_0^{\hat{t}} E(\mu_t)\psi(t) dt \geqslant \int_0^{\hat{t}} \int_{\mathbb{R}^m} \left\langle -V_t, T_{\mu_t}^\nu - Id \right\rangle \psi(t) d\mu_t dt + \frac{\lambda}{2} \int_0^{\hat{t}} W_2^2(\mu_t, \nu)\psi(t) dt. \tag{30}$$

Let t_0 be a Lebesgue point of the map $t \mapsto \int_0^{\hat{t}} E(\mu_t)\psi(t) dt + \int_{\mathbb{R}^m} \left\langle V_t, T_{\mu_t}^\nu - Id \right\rangle d\mu_t + \frac{\lambda}{2} \int_0^{\hat{t}} W_2^2(\mu_t, \nu)\psi(t) dt$. By considering a sequence of smooth mollifiers ψ_n converging to the delta function at t_0, the inequality (30) is reduced to

$$E(\nu) - E(\mu_{t_0}) \geqslant \int_{\mathbb{R}^m} \left\langle -V_{t_0}, T_{\mu_{t_0}}^\nu - Id \right\rangle d\mu_{t_0} + \frac{\lambda}{2} W_2^2(\mu_{t_0}, \nu). \tag{31}$$

Therefore (31) holds for almost all t. By Lemma 2.2 we have

$$V_t \in -\partial E(\mu_t) \quad \text{for almost all } t \in [0, \hat{t}].$$

We now study uniqueness of the solution. The available uniqueness proofs in the case of λ-convexity can be repeated in the domain of restricted λ-convexity because as soon as the flow exists clearly it dissipates the energy and the trajectories of the flow starting from μ remain in the domain of restricted λ-convexity for a short time where one can use λ-convexity. Hence uniqueness arguments can be repeated similar to the available proofs such as in [1, Theorem 11.1.4]. Therefore we provide only the key ideas here.

Assume that we have two gradient flows μ_t^1 and μ_t^2 both starting from μ. One can show that $W_2^2(\mu_t^1, \mu_t^2)$ is absolutely continuous in time and

$$\frac{d}{dt} W_2^2(\mu_t^1, \mu_t^2) \leqslant \lim_{h\downarrow 0} \frac{W_2^2(\mu_{t+h}^1, \mu_{t+h}^2) - W_2^2(\mu_{t+h}^1, \mu_t^2)}{h} + \lim_{h\downarrow 0} \frac{W_2^2(\mu_{t+h}^1, \mu_t^2) - W_2^2(\mu_t^1, \mu_t^2)}{h}.$$

By differentiability of Wasserstein metric (6), for almost all $t \in [0, \hat{t}]$ we have

$$\frac{1}{2} \lim_{h\to 0} \frac{W_2^2(\mu_{t+h}^1, \mu_t^2) - W_2^2(\mu_t^1, \mu_t^2)}{h} = \int_{\mathbb{R}^m} \left\langle V_t^1, Id - T_{\mu_t^1}^{\mu_t^2} \right\rangle d\mu_t^1.$$

16

Therefore

$$\frac{d}{dt}W_2^2(\mu_t^1, \mu_t^2) \leqslant \int_{\mathbb{R}^m} \left\langle V_t^2, Id - T_{\mu_t^2}^{\mu_t^1} \right\rangle d\mu_t^2 + \int_{\mathbb{R}^m} \left\langle V_t^1, Id - T_{\mu_t^1}^{\mu_t^2} \right\rangle d\mu_t^1. \tag{32}$$

Consider (31) along μ_t^1. We have

$$E(\mu_t^2) - E(\mu_t^1) \geqslant \int_{\mathbb{R}^m} \left\langle -V_t, T_{\mu_t^1}^{\mu_t^2} - Id \right\rangle d\mu_t^1 + \frac{\lambda}{2}W_2^2(\mu_t^1, \mu_t^2). \tag{33}$$

Rewriting (33) again along μ_t^2 and using (32) result in

$$\frac{1}{2}\frac{d}{dt}W_2^2(\mu_t^2, \mu_t^1) \leqslant -\lambda W_2^2(\mu_t^2, \mu_t^1).$$

Hence

$$W_2(\mu_t^2, \mu_t^1) \leqslant e^{-\lambda t}W_2(\mu_0^2, \mu_0^1) = 0 \quad \forall t \in [0, \hat{t}].$$

$\square$

3 Wasserstein gradient flow of the Dirichlet energy

In this section we prove a local well-posedness result for the gradient flow of the Dirichlet energy on S^1. Energy or in short E in this section always refers to the **Dirichlet energy**

$$E(\mu) = \begin{cases} \dfrac{1}{2}(\partial_x u)^2 & \text{if } \mu = udx, u \in H^1(S^1), \\ +\infty & \text{else.} \end{cases} \tag{34}$$

When convenient, we refer to an absolutely continuous measure $\mu = udx$ by its density u. In particular, by a smooth or positive measure we mean a measure with a smooth or positive density.

3.1 Dirichlet energy on S^1

The underlying space of the measures that we study in this section is S^1. We identify S^1 with $\mathbb{R}/\mathbb{Z}$. Because S^1 is a manifold, the theory developed in the previous section should be slightly modified. On a Riemannian manifold, there is the issue of existence and regularity of the optimal maps. This question has been an active area of research. Ma-Trudinger-Wang condition in [18] is a famous example that studies this issue. In the case of S^m, the problem has been addressed and positive results are available such as [15] and [17]. The results guarantee that between any pair of smooth positive measures μ, ν on S^m there exists a unique smooth optimal map T_μ^ν. To apply Theorem 2.4 to S^1, one has to replace the inner product in the subdifferential definition (9) with the inner product on the tangent space of S^1. For a given optimal map $\hat{T}$, the distance $|\hat{T}(x) - x|$ might not coincide with the geodesic distance of

S^1. As was suggested in [10], this problem can be solved by representing $T : [0, 1] \to [-\frac{1}{2}, \frac{3}{2}]$ where $T(x)$ is the smallest element of $\hat{T}(x) \in \mathbb{R}/\mathbb{Z}$ such that $|T(x) - x| \leqslant \frac{1}{2}$ and by relabelling S^1 by $[T(0), T(0) + 1]$. It is easy to check that T is an optimal map from $[0, 1]$ to $[T(0), T(0) + 1]$, furthermore T is monotone and the geodesic distance of S^1 coincides with $|T(x) - x|$.

Let $\mu_0 = u_0 dx$ and $\mu_1 = u_1 dx$ be two smooth measures on S^1 and let T be the optimal map between them. Then by Monge-Ampère equation (3) we have

$$u_1(T(X)) = \frac{u_0(x)}{T'(x)}.$$

Since the geodesics are given by the push forward of linear interpolation of the optimal map and the identity map, the explicit form of the geodesic u_s between u_0 and u_1 is given by

$$u_s((1 - s)x + sT(x)) = \frac{u_0(x)}{(1 - s) + sT'(x)}.$$

In the notation of the previous section, if we think of $f = T - Id$ as the tangent vector field that connects u_0 to u_1, the geodesic equation can be written as

$$u_s(x + sf(x)) = \frac{u_0(x)}{1 + sf'(x)}. \tag{35}$$

We will see in Lemma 3.3 that for studying restricted λ-convexity of the energy, it is enough to consider measures with smooth and positive densities. By the derivative formulation of λ-convexity (16) the energy is λ-convex along the geodesic μ_s connecting μ_0 to μ_1 if

$$\frac{d^2}{ds^2}E(u_s) \geqslant \lambda W_2^2(u_0, u_1). \tag{36}$$

We start by proving that the Dirichlet energy is not λ-convex for any λ. This is known to the community, and in [10] Carrillo and Slepčev proved that the Dirichlet energy is not convex on S^1. We will study a scalable family of functions to prove the lack of convexity for any $\lambda \in \mathbb{R}$. Let u_s be the geodesic connecting $u_0 = u$ to u_1, and let f be the corresponding tangent vector field. We compute the second derivative of the energy along the geodesic

$$\frac{d^2 E(u_s)}{ds^2}\bigg|_{s=0} = \frac{d^2}{ds^2}\bigg|_{s=0} \int_{S^1} (\partial_y u_s(y))^2 \, dy \,.$$

By the Monge-Ampère equation (35) and the change of variables $y = x + sf(x)$ we have

$$\begin{aligned}
\frac{d^2 E(u_s)}{ds^2}\bigg|_{s=0} &= \frac{d^2}{ds^2}\bigg|_{s=0} \int_{S^1} \left(\frac{\partial x}{\partial y} \frac{\partial}{\partial x} \frac{u(x)}{1 + sf'(x)} \right)^2 \frac{\partial y}{\partial x} dx \\
&= \frac{d^2}{ds^2}\bigg|_{s=0} \int_{S^1} \left(\frac{1}{1 + sf'(x)} \partial_x \frac{u(x)}{1 + sf'(x)} \right)^2 (1 + sf'(x)) \, dx \\
&= \frac{d^2}{ds^2}\bigg|_{s=0} \int_{S^1} \frac{((1 + sf'(x))u'(x) - su(x)f''(x))^2}{(1 + sf'(x))^3} \, dx \\
&= 2 \int_{S^1} \left\{ (f''u)^2 + 8(f''u)(f'u') + 6(f'u')^2 \right\} dx.
\end{aligned} \tag{37}$$

If the energy E is λ-convex, then

$$\underbrace{\int_{S^1} \left\{ (f''u)^2 + 8(f''u)(f'u') + 6(f'u')^2 \right\} dx}_{A} \geqslant \lambda \underbrace{\int_{S^1} f^2 u \, dx}_{B}. \tag{38}$$

We use (38) as a guide to find a counter-example. In the example, u and f are not smooth, and we cannot directly apply this computation but, Lemma 3.3 validates the calculations.

We view S^1 as the interval $[-1/2, 1/2]$ with the endpoints identified. The construction of the example is simple: let $u = 1 - 4|x|$ and $f' = u^{-1}$, this forces the integrand of A to be negative, and the rest follows from a scaling argument. We have to make some modifications to the functions so that the integral converges and the mass is normalized to 1. We define u and f' as follows

$$u(x) = \begin{cases} \frac{81}{16}(1 - 4|x|), & 0 \leqslant |x| \leqslant \frac{2}{9}, \\ \frac{9}{16}, & \frac{2}{9} \leqslant |x| \leqslant \frac{3}{8}, \\ \frac{9}{4}(1 - 2|x|), & \frac{3}{8} \leqslant |x| \leqslant \frac{1}{2} \end{cases} \quad f'(x) = \begin{cases} \frac{16}{81(1-4|x|)}, & 0 \leqslant |x| \leqslant \frac{2}{9}, \\ \frac{16}{11}(3 - 8|x|), & \frac{2}{9} \leqslant |x| \leqslant \frac{3}{8}, \\ 0, & \frac{3}{8} \leqslant |x| \leqslant \frac{1}{2}. \end{cases}$$
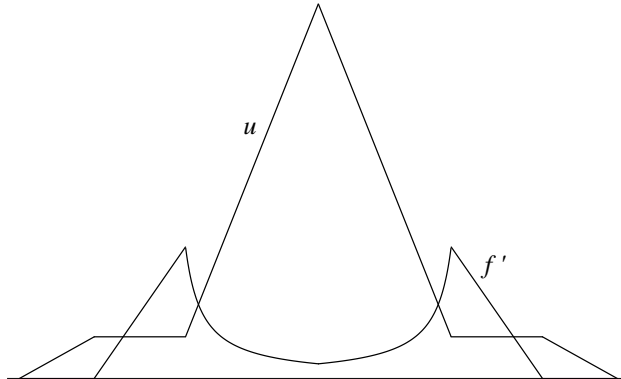

Figure 2: Graph of u and f'

By scaling $u_h(x) := hu(hx)$ and $f'_h(x) := \frac{1}{h}f'(hx)$ we have

$$A = -C_1 h^2, \quad B = C_2$$

for some positive constants C_1, C_2. If the energy is λ-convex then we must have $A \geqslant \lambda B$ and it should hold uniformly for any such u and f. But for a fixed λ, we can choose h large enough so that the opposite inequality holds. This means that the Dirichlet energy is not λ-convex on $\mathcal{P}_2(S^1)$.

In the example above, by pushing h to larger numbers the lack of convexity becomes worse. By looking at the equation of $u(x)$, it is clear that the Dirichlet energy of $u(x)$ gets bigger for larger values of h. This example hints that one of the obstructions against the λ-convexity of the Dirichlet energy is the magnitude of the energy which can be controlled on energy sub-level sets.

3.2 Restricted λ-convexity of Dirichlet energy

Lemma 3.1 (Uniform convergence on energy sub-level sets.) *Uniform convergence and Wasserstein convergence are equivalent on energy sub-level sets of Dirichlet energy on S^1. In particular for two measures $\mu_1 = u_1 dx$ and $\mu_2 = u_2 dx$ with $E(\mu_1), E(\mu_2) < c < +\infty$ we have*

$$W_2^2(\mu_1, \mu_2) \geqslant \alpha |u_1 - u_2|_\infty^\beta \tag{39}$$

where $\alpha = \alpha(c)$ and β are constants.

Proof. One side of the equivalence is easy. Assuming $u_n \xrightarrow{\ uniform\ } u_0$ we have

$$\int_{S^1} \psi u_n dx \longrightarrow \int_{S^1} \psi u dx \quad \forall \psi \in C^0(S^1)$$

which implies Wasserstein convergence of $\mu_n = u_n dx$ to $\mu = u dx$ by (8) and finiteness of the second moments on S^1.

For the converse inequality, we first study the regularity of a measure with finite energy. Let $\nu = v dx \in E_c$. By Poincare's inequality and $\int_{S^1} v dx = 1$ we have

$$\int_{S^1} |v|^2 dx \leqslant \int_{S^1} |v'|^2 dx + 2 \int_{S^1} |v| dx + \int_{S^1} dx \leqslant c + 3.$$

Therefore H^1-norm of v is bounded by its energy. The Sobolev embedding theorem implies that v is $C^{0,1/2}$ continuous and we have

$$|v(x) - v(y)| \leqslant \left(\int_{S^1} |v'|^2 dx \right)^{1/2} \left(\int_x^y dx \right)^{1/2} \leqslant \sqrt{c|x - y|} \quad \forall x, y \in S^1. \tag{40}$$
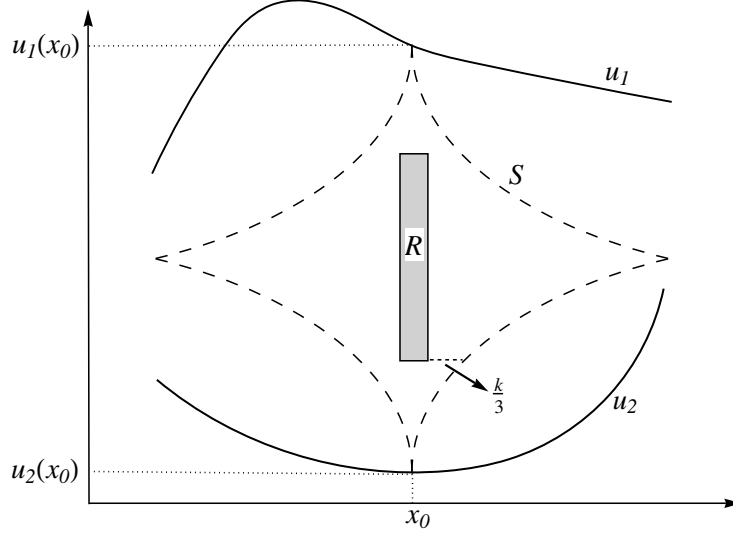
20

Figure 3: Wasserstein $\Longleftrightarrow$ Uniform

Therefore the modulus of continuity is $\sqrt{c}$. Let μ_1 and μ_2 be as in the assumption. Therefore u_1 and u_2 are $C^{0,1/2}$ continuous with constant $\sqrt{c}$. Assume that $|u_1 - u_2|_\infty \geqslant h > 0$. In particular without loss of generality assume that for some point $x_0 \in S^1$ we have $u_1(x_0) - u_2(x_0) \geqslant h$. For every $x \in S^1$, we have

$$
\begin{aligned}
u_1(x) &\geqslant u_1(x_0) - \sqrt{c|x|} \\
u_2(x) &\leqslant u_2(x_0) + \sqrt{c|x|}.
\end{aligned}
\tag{41}
$$

Therefore u_1 lies above and u_2 is below the star-like shape in Figure (3). Call the star-like shape by S. Consider a rectangle R in the center of S with height $\frac{h}{2}$ and width $\frac{k}{3}$ where k is the width of S at the height $u_2(x_0) + \frac{h}{4}$. We have $k = \frac{h^2}{8c}$ and the area of R is given by $\frac{h^3}{48c}$. In order to transport the measure μ_1 to μ_2, some mass at least equal to the area of R should be transported outside of S. Therefore

$$
\begin{aligned}
W_2^2(\mu_1, \mu_2) &\geqslant \{\text{area of R}\}\{\text{distance required to move R outside of S}\}^2 \\
&\geqslant \frac{h^3}{48c} \cdot (\frac{k}{3})^2 \\
&\geqslant \frac{1}{384c^3} |u_1 - u_2|_\infty^7.
\end{aligned}
$$

$\square$

Lemma 3.2 (Lower Semi Continuity.) *The Dirichlet energy is lower semi continuous with respect to the Wasserstein metric on S^1.*

Proof. Let $u_n \xrightarrow{W_2} u$. Since we have to prove $E(u) \leqslant \liminf_{n \to \infty} E(u_n)$, we can assume that $E(u_n)$ is bounded. This implies that H^1-norm of the sequence is bounded. By Banach–Alaoglu theorem, u_n has a weak limit point $v \in H^1$. Therefore u_n converges to v strongly in L^2. We claim that $v = u$. Let $\psi \in C^0(S^1)$ we have

$$\left| \int_{s^1} \psi(x)(u-v)(x)dx \right| \leqslant \left| \int_{s^1} \psi(x)(u-u_n)(x)dx \right| + \left| \int_{s^1} \psi(x)(u_n - v)(x)dx \right|$$

$$\leqslant \left| \int_{s^1} \psi(x)(u-u_n)(x)dx \right| + \left(\int_{s^1} |\psi|^2 dx \right)^{1/2} \left(\int_{s^1} |u_n - v|^2 dx \right)^{1/2}$$

By Lemma 3.1 the Wasserstein and uniform convergences are equivalent on energy sub-level sets. Therefore the first term in the last inequality goes to zero. The second term also goes to zero because u_n converges to v strongly in L^2. Hence $u = v$ almost everywhere. Because u and v are continuous, we have $u = v$. The Dirichlet energy is known to be lower semi continuous under weak H^1 convergence (for example see [14, Theorem 8.2.1]). Hence we have $E(u) = E(v) \leqslant \liminf_{n \to \infty} E(u_n)$. $\qquad\square$

The following lemma validates smooth calculation in the sense that for studying restricted λ-convexity of the energy, one can study restricted λ-convexity of the energy only on smooth measures.

Lemma 3.3 (Approximation by smooth measures.) *Let $\mu \in E_c$. Assume that the energy is restricted λ-convex on smooth measures in $B_\delta \cap E_c$. Then E is restricted λ-convex at μ.*

Proof. Let $\mu_0 = u_0 dx$, $\mu_1 = u_1 dx \in B_\delta \cap E_c$ and let η_k be a standard smooth mollifier converging to the Dirac delta function. Define $u_{k,i}(x) := \eta_k * u_i(x)$ for $i = 0, 1$ where $*$ is the convolution on S^1. Since $u_{k,i} \xrightarrow{uniformly} u_i$, by Lemma 3.1 we have $u_{k,i} \xrightarrow{W_2} u_i$. Therefore for large enough k we have $u_{k,i} \in B_\delta(u)$. The energy also converges, because

$$E(u_{k,i}) = \int_{S^1} (\partial_x(u_i * \eta_k))^2 \, dx = \int_{S^1} ((\partial_x u_i) * \eta_k)^2 \, dx \xrightarrow[k \to \infty]{} \int_{S^1} (\partial_x u_i)^2 \, dx = E(u_i). \quad (42)$$

Hence $u_{k,i} \in B_\delta(u) \cap E_c$ for large enough k. By smoothness of $u_{k,i}$ and the assumption of the lemma, we have λ-convexity of the energy along the geodesics $u_{k,s}$ connecting $u_{k,0}$ to $u_{k,1}$

$$E(u_{k,s}) \leqslant (1-s)E(u_{k,0}) + sE(u_{k,1}) - \frac{\lambda}{2}s(1-s)W_2^2(u_{n,0}, u_{k,1}). \quad (43)$$

Let γ_k and γ be in order the optimal plan connecting $\mu_{k,0} = u_{k,0}dx$ to $\mu_{k,1} = u_{k,1}dx$ and the optimal plan connecting μ_0 to μ_1. By stability of the optimal plans [23, Theorem 5.20] γ_k converges in narrow topology to γ along a subsequence which after relabelling we assume to

be the whole sequence. Equivalence of narrow and Wasserstein convergence (39) on $S^1 \times S^1$ implies

$$\mu_{k,s} = \left((1-s)\Pi^1 + s\Pi^2\right)_\# \gamma_k \xrightarrow{W_2} \left((1-s)\Pi^1 + s\Pi^2\right)_\# \gamma = \mu_s$$

where Π^i is the projection to the ith coordinate and μ_s is the geodesic connecting μ_0 to μ_1. The lower semi-continuity of Dirichlet energy 3.2 yields $E(\mu_s) \leqslant \liminf_{k\to+\infty} E(\mu_{k,s})$. Hence by taking the limit of (43) we have

$$E(\mu_s) \leqslant (1-s)E(\mu_0) + sE(\mu_1) - \frac{\lambda}{2}s(1-s)W_2^2(\mu_0, \mu_1).$$

$\square$

In the following lemma we prove that the energy is finite along a geodesic, provided that the energies of the end points are finite.

Lemma 3.4 (Energy of the interpolant.) *Let $\mu_0 = u_0 dx$ and $\mu_1 = u dx$ be two smooth measures with $E(\mu_0), E(\mu_1) < c < +\infty$ and $u_0, u_1 > m > 0$. Then there are constants $\hat{c} < +\infty$ and $\hat{m} > 0$ depending only on c and m such that $E(\mu_s) < \hat{c}$ and $u_s > \hat{m}$ along the geodesic μ_s connecting μ_0 to μ_1.*

Proof. By $C^{0,1/2}$ continuity of the densities, there exists $M = M(c)$ such that $u_0(x), u_1(x) < M$ for all $x \in S^1$. Let $T : S^1 \longrightarrow S^1$ be the the optimal transport map between u_0 and u_1. By Monge–Ampère equation (3) we have

$$|T'(x)| \leqslant \frac{M}{m}$$

By taking the derivative of Monge–Ampère equation (3) we have

$$\begin{aligned}
|T''(x)| &= \left| \frac{u_0'(x)u_1(T(x)) - u_0(x)T'(x)u_1'(T(x))}{u_1(x)^2} \right| \\
&\leqslant \frac{M}{m^2}|u_0'(x)| + \frac{M^2}{m^3}|u_1'(x)|.
\end{aligned} \tag{44}$$

Now let μ_s be the geodesic connecting μ_0 to μ_1. We have

$$u_s((1-s)x + sT(x)) = \frac{u_0(x)}{(1-s) + sT'(x)}. \tag{45}$$

Plugging in bounds on T' yields

$$u_s(x) \geqslant \frac{m^2}{M}.$$

Hence $u_s > \hat{m}$ where $\hat{m} = \hat{m}(c,m)$. Taking derivative of the equation (45) we have

$$u_s'((1-s)x + sT(x)) = \frac{u_0'(x)[(1-s) + sT'(x)] - su_0(x)T''(x)}{((1-s) + sT'(x))^3}$$

Using the bounds on T' and (44) we have:

$$|u_s'((1-s) + sT'(x))| \leqslant \frac{|u_0'(x)|}{|(1-s) + sT'(x)|^2} + \frac{|u_0(x)||T''(x)|}{|(1-s) + sT'(x)|^3}$$

$$\leqslant (\frac{M}{m})^2 |u_0'(x)| + (\frac{M}{m})^5 |u_0'(x)| + (\frac{M}{m})^6 |u_1'(x)|.$$

Taking integral from both sides yields $E(u_s) < \hat{c}$ where $\hat{c} = \hat{c}(c, m)$. $\qquad\square$

The idea of the following lemma was suggested by my supervisor Almut Burchard. This lemma will be used in calculations of the second derivative of the energy in Theorem 3.1.

Lemma 3.5 (Interpolation inequality.) *For every $\alpha \in \mathbb{R}$ there exists a constant $\lambda < 0$ such that*

$$|f''|_{L^2} - \alpha|f'^2|_\infty - \lambda|f|_{L^2} \geqslant 0 \quad \forall f \in C^\infty(S^1).$$

Proof. Consider the Fourier expansion $f(x) = \sum_{k \in \mathbb{Z}} a_k e^{i2k\pi x}$. We have

$$|f'|_\infty = \sup_{x \in S^1} |\sum_{k \in \mathbb{Z}} i2k\pi a_k e^{i2k\pi x}| \leqslant 2\pi \sum_{k \in \mathbb{Z}} |ka_k| = 2\pi \sum_{k \in \mathbb{Z}} |k^4 a_k^2|^{\frac{4}{10}} |a_k^2|^{\frac{1}{10}} |\frac{1}{k}|^{\frac{6}{10}}.$$

Hölder's inequality with exponents $\frac{4}{10}$, $\frac{1}{10}$, and $\frac{5}{10}$ yields

$$|f'|_\infty \leqslant 2\pi \left(\sum_{k \in \mathbb{Z}} (|k^4 a_k^2|)\right)^{\frac{4}{10}} \left(\sum_{k \in \mathbb{Z}} |a_k^2|\right)^{\frac{1}{10}} \left(\sum_{k \in \mathbb{Z}} |\frac{1}{k}|^{\frac{6}{5}}\right)^{\frac{1}{2}}.$$

The term $2\pi(\sum_{k \in \mathbb{Z}} |\frac{1}{k}|^{\frac{6}{5}})^{1/2} = d$ is a constant independent of a_k. Therefore

$$|f'|_\infty \leqslant d|f''|_{L^2}^{4/5} |f|_{L^2}^{1/5}.$$

By the arithmetic-geometric inequality for a constant β we have

$$|f'|_\infty \leqslant d|f''|_{L^2}^{4/5} |f|_{L^2}^{1/5}$$

$$= d(\beta^{5/4}|f''|_{L^2})^{4/5} (\beta^{-5}|f|_{L^2})^{1/5}$$

$$\leqslant \frac{4d}{5}\beta^{5/4}|f''|_{L^2} + \frac{d}{5}\beta^{-5}|f|_{L^2}.$$

Putting $\beta = (\frac{5}{4}\alpha d)^{\frac{-4}{5}}$ and $\lambda = -\frac{d\alpha\beta^{-5}}{5}$ yields

$$\alpha|f'^2|_\infty \leqslant |f''|_{L^2} - \lambda|f|_{L^2}.$$

$\qquad\square$

We are now ready to prove the main theorem of this section which shows that the Dirichlet energy is restricted λ-convex at positive measures.

Theorem 3.6 (restricted λ-convexity of the Dirichlet energy.) *Let $\mu = udx$ be a measure with $E(u) < c < +\infty$ and $u > m > 0$. Then $\exists \lambda = \lambda_{c,m}$ such that E is restricted λ-convex at μ.*

Proof. We first claim that the second derivative of the energy at a positive measure ν is uniformly bounded from below along any smooth vector field. Let $\nu = vdx$ be a measure with $E(v) < c$ and $v > m$. Let ν_1 be another smooth measure and let f be the vector field defining the geodesic $\nu_s = (Id + sf)_\# \nu$ that connects ν to ν_1. By (37) we have

$$\left. \frac{d^2 E(\nu_s)}{ds^2} \right|_{s=0} = 2 \int_{S^1} (f''v)^2 + 8(vf'')(v'f') + 6(v'f')^2 dx.$$

Recall that $W_2^2(\nu, \nu_1) = \int_{S^1} f(x)^2 v(x) dx$. By (36) the energy is λ-convex at v, if for all such vector fields

$$2 \int_{S^1} (vf'')^2 + 8(vf'')(v'f') + 6(v'f')^2 dx - \lambda \int_{S^1} vf^2 dx \geqslant 0.$$

By completing the squares we have

$$2 \int_{S^1} \{(vf'')^2 + 8(vf'')(v'f') + 6(v'f')^2\} dx - \lambda \int_{S^1} vf^2 dx \geqslant \int_{S^1} \{f''^2 v^2 - 52 f'^2 v'^2\} dx - \lambda \int_{S^1} vf^2 dx.$$

The lower bound on the density $v > m$ yields

$$\int_{S^1} \{f''^2 v^2 - 52 f'^2 v'^2\} dx - \lambda \int_{S^1} vf^2 dx \geqslant m^2 \int_{s^1} f''^2 dx - 52 \int_{s^1} f'^2 u'^2 dx - m\lambda \int_{s^1} f^2 dx.$$

Hölder's inequality and energy bound $E(u) < c$ imply

$$m^2 \int_{s^1} f''^2 dx - 52 \int_{s^1} f'^2 u'^2 dx - m\lambda \int_{s^1} f^2 dx \geqslant m^2 \int_{s^1} f''^2 dx - 52c|f'^2|_\infty - m\lambda \int_{s^1} f^2 dx.$$

By reordering and absorbing the constants in λ, the energy is λ-convex along ν_s at ν if $\forall f \in C^\infty(S^1)$ we have

$$|f''|_{L^2}^2 - \alpha |f'|_\infty^2 - \lambda |f|_{L^2}^2 \geqslant 0 \tag{46}$$

where $\alpha = \frac{52c}{m^2}$. By Lemma 3.5 the claim has been proved.

   Now consider the energy sub-level set E_c. By Theorem 3.1 Wasserstein convergence implies uniform convergence on E_c. Therefore there exists a $\delta = \delta_c$ such that we have $v > m$ for all $\nu = vdx \in E_c \cap B_\delta(\mu)$. Assume that $\nu_0, \nu_1 \in E_c \cap B_\delta(\nu)$. Let ν_s be the geodesic connecting ν_0 to ν_1. By Lemma 3.4 there exist $\hat{m}$ and $\hat{c}$ depending only on c and m such that $E(\nu_s) < \hat{c} < +\infty$ and $v_s > \hat{m} > 0$. By the argument at the beginning of the proof there exists a $\hat{\lambda} = \hat{\lambda}_{m,c}$ such that E is $\hat{\lambda}$-convex along the geodesic ν_s. The constant $\hat{\lambda}$ is uniform for all pairs of smooth measures inside $E_c \cap B_\delta(\mu)$. Therefore, by Lemma 3.3 E is restricted $\hat{\lambda}$-convex at μ. $\qquad\square$

Corollary 3.7 *The gradient flow trajectory of the Dirichlet energy on S^1 with a positive initial data exists and is unique at least for a short period of time.*

Corollary 3.8 *The positive periodic solutions of the thin-film equation $\partial_t u = -\partial_x(u\partial_x^3 u)$ are locally well-posed.*

4 Other classes of equations

In this section, we show that the theory developed in the last two sections can be applied to a wide class of energy functionals and evolution equations of higher order and different forms. Note that the result of Theorem 2.4 is general and it can be applied to any energy functional, provided that it is restricted λ-convex. The corresponding lemmas from Section 3 for the energies studied here can be derived in a similar fashion with minor modifications. Hence, we discuss the proofs only briefly.

4.1 Higher order equations

The family that we study here is of the form $E(u) = \frac{1}{2}\int_{S^1}|u^{(k)}|^2 dx$ for $k \in \mathbb{N}$. The flow of this family of energies corresponds to the solution of the higher order non-linear equations of the form $\partial_t u = (-1)^k \partial_x(u\partial_x^{2k+1}u)$.

Consider $u \in D(E)$. Finiteness of $|u|_{H^k}$ in particular implies that the H^1-norm of u is bounded. Since we only used the H^1-norm bounds in Lemmas 3.1, 3.2, and 3.3, they automatically follow for this class of energies. Therefore, Wasserstein and uniform convergence are equivalent on energy sub-level sets, E is lower semi continuous, and one can use approximation by smooth functions to study convexity.

In Lemma 3.4 we derived bounds on T'' by taking derivatives of the explicit formula of T' given by the Monge–Ampère equation. In the same fashion, one can find bounds on higher derivatives of the optimal map by taking more derivatives of the Monge–Ampère equation. For generalization of Lemma 3.5, we have to show that $\forall \alpha\ \exists \lambda$ such that

$$|f^{(m+1)}|_{L^2}^2 - \alpha|f^{(m)}|_\infty^2 - \lambda|f|_{L^2}^2 \geqslant 0$$

for every smooth vector field f. By induction assume that for any $\alpha_m > 0$ there exists $\lambda_m \leqslant 0$ such that

$$|f^{(m)}|_\infty^2 \leqslant \frac{1}{\alpha_m}|f^{(m+1)}|_{L^2}^2 - \lambda_m|f|_{L^2}^2. \tag{47}$$

Let $\alpha_{m+1} > 0$ be given. By applying Lemma 3.5 to $f^{(m+1)}$, there exists a $\hat{\lambda} \leqslant 0$ such that

$$|f^{(m+1)}|_\infty^2 \leqslant \frac{1}{2\alpha_{m+1}}|f^{(m+2)}|_{L^2}^2 - \hat{\lambda}|f^{(m)}|_{L^2}^2. \tag{48}$$

Put $\alpha_m = -2\hat\lambda$, by (47) there exists $\lambda_m \leqslant 0$ such that

$$|f^{(m)}|_\infty^2 \leqslant \frac{-1}{2\hat\lambda}|f^{(m+1)}|_{L^2}^2 - \lambda_m|f|_{L^2}^2.$$

$|f^{(m)}|_{L^2}^2 \leqslant |f^{(m)}|_\infty^2$ on $\mathbb{R}/\mathbb{Z}$. Therefore

$$-\hat\lambda|f^{(m)}|_{L^2}^2 \leqslant \frac{1}{2}|f^{(m+1)}|_{L^2}^2 + \hat\lambda\lambda_m|f|_{L^2}^2.$$

Plugging into (48) yields

$$|f^{(m+1)}|_\infty^2 \leqslant \frac{1}{2\alpha_{m+1}}|f^{(m+2)}|_{L^2}^2 + \frac{1}{2}|f^{(m+1)}|_{L^2}^2 + \hat\lambda\lambda_m|f|_{L^2}^2.$$

Therefore

$$|f^{(m+1)}|_\infty^2 - \frac{1}{2}|f^{(m+1)}|_{L^2}^2 \leqslant \frac{1}{2\alpha_{m+1}}|f^{(m+2)}|_{L^2}^2 + \hat\lambda\lambda_m|f|_{L^2}^2.$$

By $|f^{(m+1)}|_{L^2}^2 \leqslant |f^{(m+1)}|_\infty^2$ and by setting $\lambda_{m+1} = \frac{-1}{2}\hat\lambda\lambda_m$, we have

$$\forall \alpha_{m+1} > 0 \;\; \exists \lambda_{m+1} \leqslant 0 \quad s.t. \quad |f^{(m+1)}|_\infty^2 \geqslant \frac{1}{\alpha_{m+1}}|f^{(m+2)}|_{L^2}^2 - \lambda_{m+1}|f|_{L^2}^2. \tag{49}$$

In conclusion, all the Lemmas in the previous section can be applied to higher order energies. We now study convexity of the energies along smooth vector fields on a measure $\mu = udx$ with positive density $u > m$ and finite energy $E(u) < c < \infty$.

$$\frac{d^2}{ds^2}|_{s=0}E(u_s) = \frac{d^2}{ds^2}|_{s=0}\int_{S^1}(\partial_y^k u_s(y))^2 dy$$

$$= \frac{d^2}{ds^2}|_{s=0}\int_{S^1}\{(\frac{\partial x}{\partial y}\frac{\partial}{\partial x})^k \frac{u(x)}{1+sf'(x)}\}^2\frac{\partial y}{\partial x}dx.$$

Since we study all the different orders at the same time, we consider the general form given by a polynomial P which is determined by the order of the energy. We have

$$\frac{d^2}{ds^2}|_{s=0}E(u_s) = \int_{S^1}|uf^{(k)}|^2 + P(u, u^{(1)}, ..., u^{(k)}; f, f^{(1)}, ..., f^{(k-1)})dx$$

where P is of order at most 2 with respect to each of its entries, and the order of the derivative of each term in P is at most k. At a measure with positive density and finite energy, we have $u > m$ and $|u^{(i)}|_\infty < M$ for all $i < k$ where M depends only on $E(u)$. Also $|f^{(i)}|_{L^2} \leqslant |f^{(k-1)}|_\infty$ for all $i < k - 1$. Therefore similar to the calculation of Theorem 2.4, for positive constants β_1, β_2 we have

$$\frac{d^2}{ds^2}|_{s=0}E(u_s) \geqslant \beta_1|f^{(k)}|_{L^2}^2 - \beta_2|f^{(k-1)}|_\infty^2.$$

27

Therefore E is convex at u if we can find λ such that

$$\beta_1 |f^{(k)}|^2_{L^2} - \beta_2 |f^{(k-1)}|^2_\infty \geqslant \lambda \int_{S^1} u f^2 dx.$$

This implies that the energy is λ-convex at u because by (49) for $\alpha = \alpha_{m,c}$ there exists λ such that

$$|f^{(k)}|^2_{L^2} - \alpha |f^{(k-1)}|^2_\infty - \lambda |f|^2_{L^2} \geqslant 0 \quad \forall f \in C^\infty(S^1)$$

Hence we have proved the following theorem.

Theorem 4.1 *The energies of the form*

$$E(u) = \begin{cases} \int_{S^1} |\partial_x^k u(x)|^2 dx & \mu = u dx, \ u \in H^k(S^1), \\ +\infty & else. \end{cases}$$

are restricted λ-convex on the positive measures with finite energy. In particular, periodic gradient flow solutions of

$$\partial_t u = (-1)^k \partial_x (u \partial_x^{2k+1} u)$$

with positive initial data exist and are unique for a short time.

4.2 Different forms of equations

Consider the energies of the form $E(u) = \int_{S^1} g(u, \partial_x u) dx$. We start by calculating the second derivative of the energy along a geodesic induced by a vector field $f \in C^\infty(S^1)$.

$$\begin{aligned}
\frac{d^2}{ds^2}|_{s=0} E(u_s) &= \frac{d^2}{ds^2}|_{s=0} \int_{S^1} g(u_s(y), \partial_y u_s(y)) dy \\
&= \frac{d^2}{ds^2}|_{s=0} \int_{S^1} g\left(\frac{u(x)}{1+sf'(x)}, \frac{\partial x}{\partial y}\frac{\partial}{\partial x}(\frac{u(x)}{1+sf'(x)})\right) \frac{\partial y}{\partial x} dx \\
&= \frac{d^2}{ds^2}|_{s=0} \int_{S^1} g\left(\frac{u(x)}{1+sf'(x)}, \frac{1}{1+sf'(x)}\partial_x(\frac{u(x)}{1+sf'(x)})\right) (1+sf'(x)) dx
\end{aligned}$$

where we used the change of variable $y = x + sf(x)$. Therefore we have

$$\frac{d^2}{ds^2}|_{s=0} E(\mu_s) = \int_{s^1} \begin{bmatrix} f' & f'' \end{bmatrix} A \begin{bmatrix} f' \\ f'' \end{bmatrix} dx \qquad (50)$$

where the matrix A is given by

$$A = \begin{bmatrix} 2u'g^{(0,1)} + 4u'^2g^{(0,2)} + 4uu'g^{(1,1)} + u^2g^{(2,0)} & 2ug^{(0,1)} + 2uu'g^{(0,2)} + u^2g^{(1,1)} \\ 2ug^{(0,1)} + 2uu'g^{(0,2)} + u^2g^{(1,1)} & u^2g^{(0,2)} \end{bmatrix}$$

Note that if A is positive definite, then the energy is convex. We study the class of the form $g(u, \partial_x u) = |\partial_x(u^a)|^2$ with $a > 0$. Finiteness of the energy implies that u^a is

$C^{0,\frac{1}{2}}$ continuous with modulus of continuity smaller than the energy. Because $a > 0$, u is continuous and since $\int_{S^1} u\,dx = 1$, there exists a point x_0 with $u(x_0) = 1$. Without loss of generality we assume $x_0 = 0$. We have

$$|u(x)^a - 1| \leqslant \sqrt{c|x|} \implies u(x) \leqslant (1 + \sqrt{c|x|})^{\frac{1}{a}}.$$

Therefore there exists a uniform $M < \infty$ such that $u < M$ for all $u \in E_c$. We now briefly discuss the corresponding lemmas from Section 3.

Equivalence of Wasserstein and uniform convergence on energy sub-level sets. Let $u_2(x_0) - u_1(x_0) > h$. Then we have $u_2^a(x) \geqslant u_2^a(x_0) - \sqrt{c|x - x_0|}$ and $u_1^a(x) \leqslant u_1^a(x_0) + \sqrt{c|x - x_0|}$. Therefore the star-like shape in Lemma 3.1 should be replaced by a modified version, given by $(u_2^a(x_0) - \sqrt{c|x - x_0|})^{\frac{1}{a}}$ and $(u_1^a(x_0) + \sqrt{c|x - x_0|})^{\frac{1}{a}}$, and the rest of the proof goes similarly. Hence, we have equivalence of the Wasserstein and uniform convergence on the energy sub-level sets.

Lower semi-continuity and smooth approximation. Having Lemma 3.1 for this class of energies, the proof of Lemmas 3.2 and 3.3 can be repeated by replacing u with u^a. Hence the energy $E(u) = \int_{S^1} |\partial_x u^a|\,dx$ is lower semi continuous and one can use approximation by smooth functions.

Energy of the interpolant. Let u be bounded away from zero $u > m > 0$. When $a \geqslant 1$

$$m^{2(a-1)} \int_{S^1} |\partial_x u|^2 dx \quad \leqslant \quad E(u) = \int_{S^1} u^{2(a-1)} |\partial_x u|^2 dx \quad \leqslant \quad M^{2(a-1)} \int_{S^1} |\partial_x u|^2 dx \qquad (51)$$

and when $0 < a < 1$

$$M^{2(a-1)} \int_{S^1} |\partial_x u|^2 dx \quad \leqslant \quad E(u) = \int_{S^1} u^{2(a-1)} |\partial_x u|^2 dx \quad \leqslant \quad m^{2(a-1)} \int_{S^1} |\partial_x u|^2 dx. \qquad (52)$$

By equivalence of Wasserstein and uniform convergence, there exists δ such that $v > m$ for all $v \in B_\delta(u) \cap E_c$. Also we have proved that $v < M$ for all $v \in E_c$. Therefore we can refer to the calculation for the Dirichlet energy and just compare the energy of the geodesic with the corresponding Dirichlet energy using (51) and (52) to find a bound on the energy of interpolate points along a geodesic.

In conclusion, all of the required lemmas are true. By (50), along a geodesic induced by a smooth vector field f we have

$$\frac{d^2}{ds^2}\Big|_{s=0} E(u_s) = \int_{S^1} 2a^2 u^{2(a-1)} \left((uf'')^2 + 4(1+a)(uf'')(u'f') + (1+a)(1+2a)(u'f')^2\right) dx$$

$$\geqslant \int_{S^1} \alpha_1 (u^a f'')^2 - \alpha_2 ((u^a)'f')^2 dx$$

29

for some constants α_1, α_2. Similar to (46), we have

$$\frac{d^2}{ds^2}|_{s=0}E(u_s) - m\lambda \int_{S^1} uf^2 dx \quad \geqslant \quad |f''|^2_{L^2} - \alpha|f'|^2_\infty - \lambda|f|^2_{L^2} \quad \geqslant \quad 0$$

where the last inequality follows from Lemma 3.5. We have proved the following theorem.

Theorem 4.2 *For every $a > 0$*

$$E(u) = \begin{cases} \int_{S^1} |\partial_x u(x)^a|^2 dx & \mu = udx, \ u \in H^1(S^1) \\ +\infty & else. \end{cases}$$

is restricted λ-convex on positive measures with finite energy. In particular, periodic gradient flow solutions of

$$\partial_t u = -2a\partial_x(u\partial_x(u^{a-1}\partial_x^2 u^a))$$

with positive initial data exist and are unique for a short time.

An interesting example is the **Fisher Information**

$$E(u) = \frac{1}{2}\int_{S^1} |\partial_x u(x)^{\frac{1}{2}}|^2 dx$$

which corresponds to the **quantum drift diffusion Equation**

$$\partial_t u = -\partial_x(u\partial_x \frac{\partial_x^2 \sqrt{u}}{\sqrt{u}}).$$

Therefore we have local well-posedness of periodic solutions of the quantum drift diffusion equation with positive initial data.

Another interesting case is the limiting case $a = 0$. The corresponding energy can be written as $E(u) = \frac{1}{2}\int_{S^1} |\partial_x \log u|^2 dx$. Finiteness of the energy result in $C^{0,\frac{1}{2}}$ continuity of $\log u$. All of the lemmas can be repeated in a similar fashion for this energy. Furthermore, finiteness of the energy implies a lower bound for the measure because

$$|\log u(x) - \log(1)| \leqslant \sqrt{c|x|} \quad \implies \quad e^{-\sqrt{c}} \leqslant u(x) \leqslant e^{\sqrt{c}}. \tag{53}$$

Therefore positivity is preserved along the flow. By (50) we have

$$\frac{d^2}{ds^2}|_{s=0}E(\mu_s) = \int_{S^1} 2(f'')^2 + 4(f'')(\frac{u'}{u}f') - 2(\frac{u'}{u}f')^2 dx.$$

By 3.5 there exists λ such that

$$\frac{d^2}{ds^2}|_{s=0}E(\mu_s) \quad \geqslant \quad |f''|^2_{L^2} - 6ce^{\sqrt{c}}|f'|^2_\infty - \lambda|f|_{L^2} \quad \geqslant \quad 0.$$

Hence E is restricted λ convex at $u \in E_c$. Furthermore, since there is a uniform lower bound $e^{-\sqrt{c}}$ for all $v \in E_c$, the constant λ is uniformly bounded along the flow. Therefore, the gradient flow is globally well-posed and we have the following theorem.

Theorem 4.3 *Wasserstein gradient flow of the energy*

$$E(u) = \begin{cases} \frac{1}{2}\int_{S^1} |\partial_x \log u|^2 dx & \mu = udx, \ u \in H^1(S^1) \\ +\infty & else. \end{cases}$$

is globally well-posed. Hence the equation

$$\partial_t u = -\partial_x(u\partial_x^2 \frac{\partial_x u}{u^2})$$

with periodic boundary condition is well-posed.

Remarks. There are some simple and some more challenging directions to extend the developed method to other classes of equations. As a simple application, one can construct other classes of restricted λ-convex functionals by combining the ones already studied. For example, the solution of the energy $E(u) = \int_{S^1}\{|\partial_x u|^2 + \epsilon\frac{1}{u^2}\}dx$, which is the Dirichlet energy with a perturbation, is globally well-posed. The reason is that the second term forces the energy to remain positive. One interesting problem is the analysis of equations in higher dimensions. Our method is utilizing Sobolev embedding theorem on energy sub-level sets which is getting weaker on higher dimensions. An interesting question is whether it is possible to solve this problem with studying higher order energies.

References

[1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[2] Luigi Ambrosio, Giuseppe Savaré, and Lorenzo Zambotti. Existence and stability for Fokker-Planck equations with log-concave reference measure. *Probab. Theory Related Fields*, 145(3-4):517–564, 2009.

[3] Sebastian Andres and Max-K. von Renesse. Particle approximation of the Wasserstein diffusion. *J. Funct. Anal.*, 258(11):3879–3905, 2010.

[4] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.

[5] A. L. Bertozzi and M. Pugh. The lubrication approximation for thin viscous films: regularity and long-time behavior of weak solutions. *Comm. Pure Appl. Math.*, 49(2):85–123, 1996.

[6] Adrien Blanchet, Vincent Calvez, and José A. Carrillo. Convergence of the mass-transport steepest descent scheme for the subcritical Patlak-Keller-Segel model. *SIAM J. Numer. Anal.*, 46(2):691–721, 2008.

[7] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

[8] E. A. Carlen and W. Gangbo. Constrained steepest descent in the 2-Wasserstein metric. *Ann. of Math. (2)*, 157(3):807–846, 2003.

[9] José A. Carrillo, Robert J. McCann, and Cédric Villani. Contractions in the 2-Wasserstein length space and thermalization of granular media. *Arch. Ration. Mech. Anal.*, 179(2):217–263, 2006.

[10] José A. Carrillo and Dejan Slepčev. Example of a displacement convex functional of first order. *Calc. Var. Partial Differential Equations*, 36(4):547–564, 2009.

[11] Fausto Cavalli and Giovanni Naldi. A Wasserstein approach to the numerical solution of the one-dimensional Cahn-Hilliard equation. *Kinet. Relat. Models*, 3(1):123–142, 2010.

[12] Ennio De Giorgi. New problems on minimizing movements. In *Boundary value problems for partial differential equations and applications*, volume 29 of *RMA Res. Notes Appl. Math.*, pages 81–98. Masson, Paris, 1993.

[13] Bertram Düring, Daniel Matthes, and Josipa Pina Milišić. A gradient flow scheme for nonlinear fourth order equations. *Discrete Contin. Dyn. Syst. Ser. B*, 14(3):935–959, 2010.

[14] Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.

[15] Alessio Figalli, Young-Heon Kim, and Robert McCann. Regularity of optimal transport maps on multiple products of spheres. *To appear in J. Eur. Math. Soc. (JEMS)*.

[16] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1), 1998.

[17] Grégoire Loeper. Regularity of optimal maps on the sphere: the quadratic cost and the reflector antenna. *Arch. Ration. Mech. Anal.*, 199(1):269–289, 2011.

[18] Xi-Nan Ma, Neil S. Trudinger, and Xu-Jia Wang. Regularity of potential functions of the optimal transportation problem. *Arch. Ration. Mech. Anal.*, 177(2):151–183, 2005.

[19] Robert John McCann. *A convexity theory for interacting gases and equilibrium crystals.* ProQuest LLC, Ann Arbor, MI, 1994. Thesis (Ph.D.)–Princeton University.

[20] Luca Natile and Giuseppe Savaré. A Wasserstein approach to the one-dimensional sticky particle system. *SIAM J. Math. Anal.*, 41(4):1340–1365, 2009.

[21] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

[22] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2003.

[23] Cédric Villani. *Optimal transport, Old and New*, volume 338 of *Grundlehren der Mathematischen Wissenschaften.* Springer-Verlag, Berlin, 2009.